Docquier, Frédéric; Vasilakis, Chrysovalantis; Munsi, D. Tamfutu

Working Paper

International Migration and the Propagation of HIV in Sub- Saharan Africa

FERDI Working Paper, No. P89

Provided in Cooperation with:
Fondation pour les études et recherches sur le développement international (FERDI), Clermont-Ferrand

Suggested Citation: Docquier, Frédéric; Vasilakis, Chrysovalantis; Munsi, D. Tamfutu (2014) : International Migration and the Propagation of HIV in Sub- Saharan Africa, FERDI Working Paper, No. P89, Fondation pour les études et recherches sur le développement international (FERDI), Clermont-Ferrand

This Version is available at:
https://hdl.handle.net/10419/269370

FONDATION POUR LES ÉTUDES ET RECHERCHES SUR LE DÉVELOPPEMENT INTERNATIONAL

document de travail 89 February 2014 Politiques de développement

International Migration and the Propagation of HIV in Sub-Saharan Africa*

Frédéric Docquier / Chrysovalantis Vasilakis / D. Tamfutu Munsi

Frédéric DOCQUIER is Professor of Economics at IRES, Université Catholique de Louvain. He is Research Associate at FNRS and Senior Fellow Ferdi
mail : frederic.docquier@uclouvain.be

Chrysovalantis VASILAKIS, IRES, Université Catholique de Louvain and University of Warwick, and Bangor Business School

D. TAMFUTU MUNSI, IRES, Université Catholique de Louvain.

Abstract

In this paper, we identify and quantify the role of international migration in the propagation of HIV across sub-Saharan African countries. We use panel data on bilateral migration flows and HIV prevalence rates covering 44 countries after 1990. Controlling for unobserved heterogeneity, reverse causality, reflection issues, incorrect treatment of country fixed effects and spatial auto-correlation, we find evidence of a highly robust emigration-induced propagation mechanism. On the contrary, immigration has no significant effect. Numerical experiments reveal that the long-run effect of emigration accounts for more than 4 percent of the number of HIV cases in 15 countries (and more than 20 percent in 6 countries).

JEL Classification : F22, I12, J61

Keywords : international migration, labor mobility, HIV/AIDS, pandemics, propagation of diseases.

* We thank three anonymous referees and the Editor in charge of this paper for constructive and helpful suggestions. Comments from Michel Beine, Alok Bhargava, J. Paul Elhorst, Catherine Guirkinger, Hillel Rapoport, Fatemeh Shadman-Metha, Tobias Müller and Vincenzo Verardi were also appreciated. The authors are grateful for the financial support from the Belgian French-speaking Community (convention ARC 09/14-019 on "Geographical Mobility of Workers and Firms"). The usual disclaimers apply.

1 Introduction

It has long been recognized that international migration is a powerful force that shapes the distribution of human populations across the globe. It affects economic inequality between nations and contributes to propagate economic shocks (see Hatton and Williamson, 2008; Massey, 1988; Docquier and Rapoport, 2012; Gibson and McKenzie, 2011). A recent strand of literature gives support to the view that migration also induces important transfers of political, cultural, sociological or behavioral norms and values between countries. Spilimbergo (2009) shows that foreign-educated individuals promote democracy in their home country, but only if foreign education is acquired in democratic countries. Lodigiani and Salomone (2012) demonstrate that emigration to democratic countries improves female political empowerment in the origin country. Fargues (2007), Beine et al. (2013) and Bertoli and Marchetta (2013) provide evidence of migration-induced transfers of fertility norms, i.e. fertility behavior at origin is affected by fertility rates at destination.

While the literature has mainly focused on transfers of positive norms, it is pretty obvious that movements of people can also propagate negative shocks across countries. In particular, migration is a source of propagation of pandemic diseases within and across regions. History shows that colonization served to propagate germs across countries and continents. Migration contributed to spread bubonic plague within Europe in the 14th century. It propagated the Spanish flu from East Asia to Russia, Europe and North America in the beginning of the 20th century (Diamond, 1997).

Not surprisingly, migration is also perceived as a factor explaining the spreading of HIV/AIDS within and across countries. Africa is the most infected continent with average HIV prevalence rates as high as 25 percent in the Southern and Eastern regions. The HIV virus causes AIDS, which is expected to induce the death of about 100 million people per year by 2025. Many case studies have highlighted the mechanism through which workers' mobility contributes to propagate the disease (see among others Anarfi, 1993; Decosas et al, 1995; Hope, 2001; Ateka, 2001; Brummer, 2002). Although many migrants have regular sexual partners, some have relations with casual partners and face a higher risk to be infected (Brockerhoff and Biddlecom, 1999). This is especially the case for male workers migrating or commuting to find jobs on plantations or in mines, where prostitutes are brought in. The circular nature of migration and the maintenance of links with home through frequent visits puts people at risk at both ends of the migratory movement.[1] The goal of our paper is to shed light on the relationship between international migration and the propagation of HIV in sub-Saharan Africa using macro data.

Although many cross-country studies have investigated the links between macroeconomic variables and health (e.g. Owen and Wu, 2007; Bhargava et al., 2001; Pritchett and Summers, 1996), only a few of them have analyzed the macro deter-

[1] Another factor relates to the migration of unhealthy widows away from their deceased spouse (Ntozi, 1997).

minants of HIV incidence. A noticeable exception is Oster (2012), who estimated the relationship between exports and the incidence of HIV in sub-Saharan Africa. She found a significant and large positive effect of trade: a doubling of exports leads to approximately a doubling in new HIV infections, an effect seen as resulting from increased movements of people (specifically, trucking).

Due to lack of comparable data on international migration, the role of migration has only been addressed in country-specific case studies. To the best of our knowledge, our paper is the first to quantify the effect of international migration, relying on bilateral information and standard, albeit rigorous econometric techniques. We use macrodata despite their limits (lack of infra-geographical information, imperfect measurement of migration flows and HIV prevalence rates, difficulty to interpret the mechanisms at work, etc.). With macrodata, unobserved heterogeneity is key, causation is harder to establish, and it is difficult to identify the channels of transmission. However, macrodata have also some advantages: they are comparable across countries, constructed by the same authors or institutions for different periods and countries, they cover longer horizons. In addition, beyond the mere advantage of using more observations, availability of panel data allows solving some of the problems listed above and limits the risk of misspecification and endogeneity biases (Islam, 1995 and 2003; Caselli et al., 1996; Roodman, 2009; Bazzi and Clemens, 2013).

This paper identifies and quantifies the effect of international migration on HIV spreading across sub-Saharan African countries. We take advantage of a new database on bilateral migration between sub-Saharan African countries and combine it with annual panel data on HIV prevalence rates. Our data cover 44 sub-Saharan African countries. We first use standard cross-country OLS regressions (Ordinary Least Squares) to estimate the effect of immigration and emigration on the dynamics of HIV prevalence rates after 1990. These regressions reveal that emigration to high-prevalence destination countries increases infection rates at origin. On the contrary, immigration does not generate significant effects.

The OLS technique is likely to generate inconsistent estimates because of endogeneity issues and problems related to small sample size. For this reason, we use 2SLS regressions (Two-Stage Least Squares) to solve for endogeneity; and we also use panel regressions to control for omitted variables (using country-specific fixed effects), spatial correlation and endogeneity problems. In the panel setting, we annualize bilateral migration data (observable every ten years) but fully exploit the time series dimension of HIV prevalence data. Again, our 2SLS and panel analyses confirm a significant effect of emigration to high-prevalence destinations, and no impact through immigration. When annual data are used, the emigration effect is very robust and the magnitude of the elasticity is stable across specifications. Although other mechanisms are plausible, our results are consistent with the widespread view that migrants have unprotected relations with prostitutes who were already infected in the host country. Hence, immigration does not induce significant changes in prevalence rates at destination (as in Wilson, 2012). However new migrants who have

been infected, propagate the virus to their origin countries through circulation, visits and/or return migration.

Our model explains well the evolution of HIV over the nineties. Our data show that in this period, average levels of HIV at destination decreased in 20 sub-Saharan African countries, and increased in 24 countries. These variations can be due to changes in emigration flows and/or emigrants' location choices. Numerical experiments based on our estimated parameters reveal that the effect of recent emigration flows is rather low in about half of the countries included in the sample. However the long-run effect of emigration accounts for more than 4 percent of HIV cases in 15 countries, and more than 20 percent in 6 countries. On the one hand, HIV prevalence rates in the year 2000 would have been at least 10 percent larger without decreasing emigration flows from countries such as Mauritius, Lesotho, Swaziland, Botswana, Namibia and Rwanda. On the other hand, prevalence rates would have been at least 10 percent lower without increasing emigration flows from countries such as Burkina Faso, Comoros, Liberia or Equatorial Guinea.

The remainder of this paper is organized as following. Section 2 describes the empirical model and discusses econometric issues. Data are presented in Section 3. Section 4 provides empirical results. Finally, Section 5 concludes.

2 Model

Our goal is to analyze the determinants of HIV prevalence rates, defined as the percentage of people aged 15-49 who are infected with HIV. The HIV prevalence of country i $(i = 1, ..., N)$ at year t $(t = 1, ..., T)$ is expressed as percent of the 15-49 population and denoted by $H_{i,t}$. In 2000, it ranged from about 0 percent in Mauritius to 28.6 percent in Botswana, with an average value of 7 percent in sub-Saharan Africa (see Table 1 in Section 3). In this section, we present the specification used in our empirical analysis and then discuss some econometric issues.

2.1 Benchmark specification

Our model combines the time series dimension and the cross section variation of the data. Given its stock nature, HIV prevalence rates exhibit some inertia and we need a dynamic regression model to explain their evolution. Epidemiological models have long been used to characterize the progress of epidemics (see Kermack and McKendrick, 1927). The dynamics of infection rates are traditionally modelled as a function of the product of infected by non-infected shares of the population. This model may induce cyclical properties which are not supported by the data (see Figure 2.a below) and performs badly when matching the dynamics of HIV prevalence, as

shown in Appendix C.[2]

Our preferred model is a standard β-convergence specification which features the annual log-change in HIV prevalence as the dependent variable. The explanatory variables are: past level of HIV prevalence, average level of HIV prevalence in destination countries of native emigrants from country i (denoted by $Z^e_{i,t}$), average level of HIV prevalence in origin countries of foreign immigrants to country i (denoted by $Z^i_{i,t}$). The basic specification writes as follows:

$$\begin{aligned}\Delta \ln(1+H_{i,t}) &= \alpha + \beta \ln(1+H_{i,t-1}) + \gamma \ln\left(1+Z^e_{i,t-1}\right) \\ &\quad +\delta \ln\left(1+Z^i_{i,t-1}\right) + \eta X_{i,t-1} + \varepsilon_{i,t}\end{aligned} \tag{1}$$

where $\Delta \ln(1+H_{i,t}) \equiv \ln(1+H_{i,t}) - \ln(1+H_{i,t-1})$, $Z^e_{i,t-1}$ and $Z^i_{i,t-1}$ are the average levels of HIV prevalence in emigration and immigration countries (defined below), $X_{i,t-1}$ is a set of other determinants of HIV, $(\alpha, \beta, \gamma, \delta, \eta)'$ is the vector of parameters to be estimated, and $\varepsilon_{i,t}$ is the error term.

The HIV prevalence rate is expressed as percentage of the population at risk and is defined on [0;100]. We use a specification with $\ln(1+x)$ to avoid losing observations with $x \simeq 0$, i.e. to be consistent with countries where HIV prevalence rates (domestic, at destination or at origin) are null or very small. In 1990, 18 countries (out of 44) exhibited prevalence rates smaller than 0.5 percent and four countries were below 0.1 percent (Comoros, Madagascar, Mauritius and Senegal). As time passes by, the $\ln(1+x)$ transformation becomes less important: by the year 2000, prevalence rates take large values in the majority of countries (ranging from 0 to 30) but remain low in the four countries listed above. The $\ln(1+x)$ transformation drastically improves the performance of the model over the nineties, especially at the beginning of the period.[3]

In equation (1), $Z^e_{i,t-1}$ and $Z^i_{i,t-1}$, are constructed in line with previous studies on migration-induced transfers of norms and values (see Spilimbergo, 2009; Beine et al, 2013; Lodigiani and Salomone, 2012). For $Z^e_{i,t-1}$, we add up HIV prevalence rates in destination countries of native emigrants from country i, weighted by bilateral emigration rates. The latter is defined as the ratio of emigration flow from i to j to the native population in country i. We consider migration flows (rather than stocks) to eliminate earlier migrants who settled in the destination country a long time ago (possibly before the rise of HIV) or who migrated as children. Using migration flows, we focus on recent migrants who are more likely to keep strong ties with their home country.[4] This gives:

[2] In Table A.1, we report estimates obtained with an epidemiological model and show that neither the product of infected by non-infected people nor the effect of cross-border migration are significant.

[3] In Appendix C, we provide estimates obtained without the ln(1+x) transformation and shows that our results are qualitatively robust to the transformation (see Table A.1).

[4] Our estimates strongly support the choice of migration flows. Very significant and robust results were obtained with migration flows. Less robust results available upon request were obtained with weights based on migration stocks.

$$Z^e_{i,t} \equiv \frac{1}{N_{i,t}} \sum_j M_{ij,t} H_{j,t} \tag{2}$$

where $M_{ij,t}$ stands for the emigration flow from country i to country j at time t, and $N_{i,t}$ is the resident population in country i. Our rationale is that recent emigrants maintain ties with their home country through frequent visits, especially when they migrate for seeking jobs abroad. This puts people at risk in the origin country, in line with the literature described in the introduction.

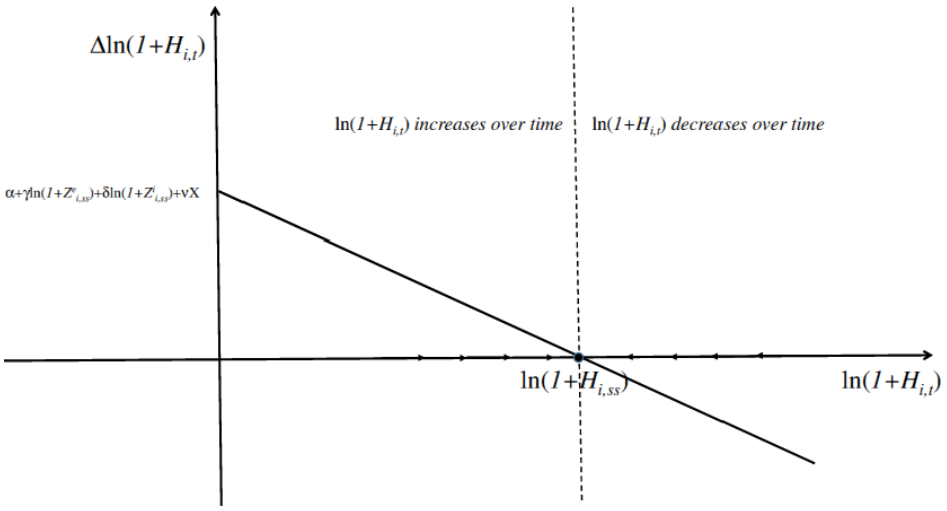
Similarly, $Z^i_{i,t}$ is the sum of HIV prevalence rates in origin countries of foreign immigrants to country i, weighted by bilateral immigration rates (defined as the ratio of bilateral immigration flow to native population):

$$Z^i_{i,t} \equiv \frac{1}{N_{i,t}} \sum_j M_{ji,t} H_{j,t} \tag{3}$$

The β-convergence model (1) has been extensively used to explain the dynamics of sluggish variables such as the stock of physical/human capital, GDP per capita, quality of institutions, etc. Figure 1 provides a graphical illustration of its dynamic properties when explanatory variables have reached their long-run levels ($Z^e_{i,t} = Z^e_{i,ss}, Z^i_{i,t} = Z^i_{i,ss}, X_{i,t} = X_{i,ss} \forall t$, and subscript ss stands for steady state). If coefficient β is negative, the growth rate of the HIV prevalence rate between years $t-1$ and t decreases linearly with its lagged level. The intercept of this linear relationship is given by $\alpha + \gamma \ln\left(1 + Z^e_{i,ss}\right) + \delta \ln\left(1 + Z^i_{i,ss}\right) + \eta X_{i,ss}$. For initial level below $\ln(1 + H_{i,ss})$, the growth rate of HIV is positive and $\ln(1 + H_{i,t})$ increases over time; for initial level above $\ln(1 + H_{i,ss})$, the growth rate of HIV is negative and $\ln(1 + H_{i,t})$ decreases over time. Hence, $\beta \in [-1; 0]$ implies that the prevalence rate gradually and monotonically converges to a stationary level in the long-run. Coefficient β determines the speed of convergence of HIV prevalence rates to their steady state level, and $\ln(1 + H_{i,ss})$ characterizes the long-run equilibrium.[5]

[5]Note that a positive value for β would imply that the growth rate of HIV prevalence increases with the current rate, i.e. a pattern of explosive dynamics. And $\beta < -1$ would imply cyclical convergence.

Figure 1. Graphical interpretation of the β-convergence model



Because our estimates will strongly support $\beta \in [-1; 0]$, we can focus on the conditional-convergence interpretation of the model: a variation in the intercept (which might be driven by emigration- or immigration-induced transfers of norms) shifts the line upwards or downwards and gives rise to a change in the long-run prevalence rate. Model (1) thus predicts that the HIV prevalence rate of country i will converge towards a long-run equilibrium level defined as

$$\ln\left(1 + H_{i,ss}\right) = \frac{\alpha + \gamma \ln\left(1 + Z^{e}_{i,ss}\right) + \delta \ln\left(1 + Z^{i}_{i,ss}\right) + \eta X_{i,ss}}{-\beta}. \tag{4}$$

Coefficient γ captures the short-run effect of emigration on HIV prevalence. If γ is positive and significant, it means that emigration to countries with high HIV prevalence rates increases the prevalence rate at origin. Coefficient δ captures the short-run effect of immigration on HIV prevalence. If δ is positive and significant, it means that immigration from countries with high HIV prevalence rates increases the prevalence rate at destination. From (4), the long-run effects of emigration and immigration are given by -γ/β and -δ/β.

2.2 Econometric issues

We first estimate Eq. (1) using standard OLS cross-country regressions. However the estimation of (1) entails several econometric issues that might lead the OLS technique to generate inconsistent estimates: omitted variables, reverse causality, correlated individual effects, spatial correlation. We address these issues by instrumenting explanatory variables, including fixed effects and using spatial regressions.

Omitted variables. The dynamics of HIV prevalence is clearly an endogenous process affected by a large number of determinants of a varied nature, captured by the vector of controls $X_{i,t}$ in equation (1). Demographic variables (age and gender structures, density, urbanization, etc.), economic variables (level of development, education, gender inequality, unemployment, structure of industry, etc.), quality of institutions (information about HIV risk, medical staffing and infrastructure, etc.) and cultural characteristics (religion and beliefs, sex practices, ethnic fractionalization, etc.) are among the main determinants. It is difficult to control for all these characteristics given the lack of data on sub-Saharan African countries. Hence, omitted variables can cause biased parameter estimates (Islam, 1995). We will take advantage of the panel dimension of our database and use country and time fixed effects. The introduction of fixed effects accounts for the time-invariant unobservable factors and common trends. Although some determinants can vary across years and countries, using fixed effects is much less restrictive that it seems at first glance. First, a lot of factors such as religion and beliefs, ethnic diversity or degree of urbanization are stable over time. Second, other factors such as education or the quality of institutions exhibit a lot of inertia. It is thus unclear whether their explicit inclusion (should we have observations for these factors) in the regression model would significantly improve the quality of fit and would reduce the degree of misspecification bias. We hypothesize that fixed effects are informative enough to account for unobserved heterogeneity and write:

$$\eta X_{i,t-1} = \alpha_i + \alpha_t$$

where α_i is the fixed effect for country i and α_t is the fixed effect for year t. We will first estimate the model with fixed effects using the Least Square Dummy Variable method (referred to as LSDV).

Other endogeneity issues. The OLS and LSDV regression models assume that all covariates are independent of the error term. In a panel setting, fixed effects control for possible misspecifications caused by unobserved characteristics; however they do not account for other possible sources of endogeneity of the regressors. Endogeneity problems may arise for several reasons. First, a positive effect of emigration or immigration on HIV prevalence could be explained by reverse causality if migration rates and destination choices are endogenous. For instance, people originating from a risky country could be willing to emigrate more. A second endogeneity source comes from the reflection problem (Manski, 1993). If country-specific equations were written as a system, the HIV prevalence rate in country i would depend on that in country j, which itself depends on that of country i. It follows that the model might not be identified, and/or suffers from a simultaneity issue, at least if both $M_{ji,t}$ and $M_{ij,t}$ are positive.[6] Third, equation (1) is dynamic given the presence of $\ln H_{i,t-1}$

[6] Concerning the identification problem we are in line with Calvo-Armengol, Patacchini and Zenou (2009) who show that network models are identified if and only if networks of individuals are not similar. The weights in the social connections in the network context play the same role as the migration structure in our context. However, the simultaneity issue remains important.

on the right-hand side. The use of fixed effects and AR terms leads to inconsistency of estimates (Nickell, 1981). Although the ratio of cross-section to time dimensions suggests that the Nickell bias should be limited in our regressions, it is interesting to look at alternative approaches.

To address reverse causality issues, we use Two-Stage Least Squares regressions (referred to as 2SLS) with external instrumental variables. We first predict bilateral migration flows using a pseudo-gravity model with exogenous determinants (a dummy equal to one if countries shared the same colonizer, a dummy equal to one if they share the same language, a dummy equal to one if they were previously the same country, and the log of geographic distance) and a full set of country and year fixed effects (Freyer, 2009). Due to the presence of a large number of zeroes or undefined observations in the dependent variable (12.5 percent of the 1,936 observations), we follow Santos Silva and Tenreyro (2006) and use the Pseudo-Poisson Maximum Likelihood (PPML) estimator. Results for the gravity regression are presented in Appendix B. Then, we compute the average HIV prevalence rates at destination and origin of migrants using the predicted (rather than observed) migration flows in (2) and (3). The resulting levels obtained for $\widehat{Z}^e_{i,t}$ and $\widehat{Z}^i_{i,t}$ are used as instruments for $Z^e_{i,t}$ and $Z^i_{i,t}$ in our first-stage regressions. It is worth noticing that the 2SLS method can be used in the cross-country setting without fixed effects or in the panel setting with fixed effects.

To better account for the the dynamic structure of our model (Islam, 2003) and for the possible endogeneity of average levels of HIV at destination and origin of migrants, we also use internal instruments. We instrument $\ln H_{i,t-1}$ and the average HIV prevalence rates at origin and destination, $\ln\left(1+Z^e_{i,t-1}\right)$ and $\ln\left(1+Z^i_{i,t-1}\right)$, with their lagged values. No need to say that these techniques can only be implemented in the panel setting. As argued by Islam (2003), there is no optimal estimation method for convergence equations in a panel data set-up. We will use the $t-2$ and $t-3$ levels to instrument $\ln\left(1+H_{i,t-1}\right)$, $\ln\left(1+Z^e_{i,t-1}\right)$ and $\ln\left(1+Z^i_{i,t-1}\right)$.

IV with differences. As argued by Caselli et al. (1996), the overwhelming majority of empirical studies on convergence are plagued by the incorrect treatment of country fixed effects. It is usually assumed that those effects are uncorrelated with the other right-hand-side variables. The fixed effect α_i in equation (1) is used as a determinant of the log-change in the HIV prevalence rate, or equivalently of $\ln\left(1+H_{i,t}\right)$. By construction, it is also a determinant of $\ln\left(1+H_{i,t-1}\right)$, which is a regressor in equation (1). Hence, the assumption of uncorrelated fixed effects is violated in panel dynamic regressions. Although fixed effects are used as control variables, it is desirable to correct for this collinearity bias. To solve this problem, Caselli et al. (1996) suggest to estimate the model in differences to eliminate country fixed effects. Equation (1) can be rewritten as

$$\begin{aligned}\ln(1+H_{i,t}) &= \alpha+(1+\beta)\ln(1+H_{i,t-1})+\gamma\ln(1+Z^e_{i,t-1})\\ &\quad +\delta\ln(1+Z^i_{i,t-1})+\alpha_i+\alpha_t+\varepsilon_{i,t}\end{aligned}$$

Differentiating yields

$$\begin{aligned}\Delta \ln(1+H_{i,t}) &= (1+\beta)\Delta \ln(1+H_{i,t-1}) + \gamma \Delta \ln(1+Z^e_{i,t-1}) \\ &\quad +\delta \Delta \ln(1+Z^i_{i,t-1}) + \widetilde{\alpha}_t + \widetilde{\varepsilon}_{i,t}\end{aligned} \tag{5}$$

where country fixed effects are eliminated, $\widetilde{\alpha}_t \equiv \alpha_t - \alpha_{t-1}$ is the new time fixed effect, and $\widetilde{\varepsilon}_{i,t} \equiv \varepsilon_{i,t} - \varepsilon_{i,t-1}$ is the transformed error term.

As above, the model in differences can be estimated after instrumenting right-hand side variables using their lagged values or using the GMM approach (Generalized Method of Moments) described in Bond (2002). We will present here the GMM results and instrument $\Delta \ln(1+H_{i,t-1})$, $\Delta \ln\left(1+Z^e_{i,t-1}\right)$ and $\Delta \ln\left(1+Z^i_{i,t-1}\right)$ with the levels observed in $t-3$ and/or $t-4$.[7]

Spatial correlation. International migration might not be the only spreading channel of HIV across nations. Commuting, tourism, trade and trucking, visits abroad, unrecorded movements of people can also propagate the virus across countries (see Oster, 2012). The size of these alternative propagation channels is reasonably linked to bilateral distance between countries. To control for this, we depart form standard LSDV and estimate two alternative models, the Spatial Error Model (SEM) and the dynamic Spatial AutoRegressive model (SAR).

In the SEM case, the spatial influence operates through the error term[8]. We estimate the SEM model assuming that the error term $\varepsilon_{i,t}$ in equation (1) exhibits spatial correlation. Like migration, the magnitude of alternative propagation channels is likely to vary with the geographic distance between countries. Hence, our $N \times N$ weight matrix W includes bilateral geographic distances between the 44 countries in our sample. Removing the country index i and using vectorial notations, we rewrite the error terms in equation (1) in the following fashion:

$$\varepsilon_t = \rho_1 W u_t + v_t \tag{6}$$

where W is the weight matrix of bilateral geographic distances between countries (with zeroes on the diagonal), ρ_1 is the spatial autoregressive parameter to be estimated, u and v are assumed to be normally and independently distributed with zero mean and constant variance.

In the SAR case, we follow the dynamic spatial autoregressive specification described in Lee and Yu (2010). We add two terms to equation (1). First, we allow the variation of HIV prevalence rate of a given country to depend on the lagged HIV prevalence rates in the other countries in addition to its own HIV prevalence rate. The difference with Z^e_{t-1} and Z^i_{t-1} is that these HIV prevalence rates abroad

[7] Similar results were obtained with a standard 2SLS method. They can be found in a previous version of this paper (See Docquier et al. 2011).

[8] For the estimation of SEM model, we use the Matlab routines for spatial panel data described in Elhorst (2003, 2010a, 2010b)

are weighted by the distance matrix W (rather than the emigration and immigration rates).[9] Coefficient λ captures the existence of alternative dynamic propagation mechanisms related to geographic distance. Second, we allow for contemporaneous spatial and social interactions between countries. We multiply the vector of current prevalence rates by the same distance matrix. Coefficient ρ_2 captures contemporaneous interactions. We estimate the model using the Quasi-Maximum Likelihood (QML) estimator as in Lee and Yu (2010).

Using vector notations (with α_n standing for the vector of country fixed effects), our general SAR specification writes as

$$\begin{aligned}\Delta \ln(1+H_t) &= \alpha + \rho_2 W \ln(1+H_t) + \lambda W \ln(1+H_{t-1}) + \beta \ln(1+H_{t-1}) \\ &\quad + \gamma \ln\left(1+Z^e_{t-1}\right) + \delta \ln\left(1+Z^i_{t-1}\right) + \alpha_n + \alpha_t + \varepsilon_t.\end{aligned} \tag{7}$$

Under the latter specification, we allow spillover effects to operate with distance (on top of migration flows to high-infection countries). This model becomes non linear in the parameters. We need to express its reduced form to obtain the spillover effects as follows:

$$\begin{aligned}(I-\rho_2 W)\ln(1+H_t) &= \alpha + [(1+\beta)I_n + \lambda W]\ln(1+H_{t-1}) \\ &\quad + \gamma \ln\left(1+Z^e_{t-1}\right) + \delta \ln\left(1+Z^i_{t-1}\right) + \alpha_n + \alpha_t + \varepsilon_t\end{aligned}$$

where I is the identity matrix. This model allows quantifying the effect of a change in the average HIV prevalence levels in emigration and immigration of a given country on its own HIV prevalence rate but also the HIV prevalence rate of all other countries in the sample. We will multiply the left and right-hand sides of the latter equation by $(I-\rho_2 W)^{-1}$, and then estimate the non linear model. We can compute the matrix of partial short-run effects of emigration and immigration as:

$$\gamma^+ = \frac{\partial \ln(1+H_t)}{\partial \ln\left(1+Z^e_{t-1}\right)} = (I-\rho_2 W)^{-1}\gamma$$

and

$$\delta^+ = \frac{\partial \ln(1+H_t)}{\partial \ln\left(1+Z^i_{t-1}\right)} = (I-\rho_2 W)^{-1}\delta$$

Off-diagonal elements of the W matrix represent indirect effects. The parameter ρ_2 measures the strength of contemporaneous spatial interdependencies between countries. If there is no contemporaneous spatial correlation, $\rho_2 = 0$, the direct effects of emigration and immigration are simply captured by γ and δ. This does not prevent the existence of alternative contagion effects if λ is positive and significant.

[9] Lee and Yu (2010) describe two approaches for the dynamic spatial model: transformation *versus* direct approach. We use the transformation approach. MATLAB codes are available upon request.

3 Data

Our sample is restricted to the 44 sub-Saharan African countries. We choose these countries because Africa is the most infected continent and HIV prevalence rates drastically increased over the nineties. We mainly focus on the period 1990-2000 given the availability of migration data but also produce results for the periods 1990-2010 and 2000-2010.

Comprehensive longitudinal data on HIV prevalence rates ($H_{i,t}$), defined as the percentage of people aged 15-49 who are infected with HIV (UNAIDS, 2008), were revised by UNAIDS for the period for 1990–2007. We also use non-revised data for the period 2008-2010; they are taken from the websites of UNAIDS and the World Health Organization. The data reveal increasing levels of HIV prevalence rates in many countries between 1991 and 2000, and large difference across countries and periods (see Figure 2.a). The same data were used in Bhargava and Docquier (2008) who study the links between HIV prevalence, medical brain drain and number of deaths due to AIDS in Africa. In Appendix A, we show that HIV prevalence rates have been much more stable since 2000.

Data on international migration are taken from Ozden et al. (2011). They collected bilateral data on migration stocks ($S_{ij,t}$) for more than 200 countries from 1960 to 2000, with one observation every ten years. Following the United Nations definition, they define a migrant as "any person that changes his or her country of usual residence" (United Nations, 2009) and classify migrants by country of birth. Compared to the United Nations' database, an important feature of the matrices presented in Ozden et al (2011) is that refugees are not accounted for.

Ozden et al. (2011) used various sources to record migrants, mainly census and population register records collected in the destination countries. They had to deal with inevitable gaps in the data. They obtained census data from 33 sub-Saharan African countries in 1990 (out of 44 in our sample), representing 90.2 percent of the total stock of immigrants. For the missing countries, they interpolated the 1990 stock using data for the earlier and later periods. As for the 2000 census round, they obtained data for 20 countries only. For the missing countries, they used the total stocks of immigrants reported in the Trends in International Migrant Stock of the United Nations (2009) and assumed these stocks have the same bilateral composition as in the previous decade. We believe their imputation strategy for 2000 will not drive or distort our results for three reasons. First, the bilateral structure of migration stock is a very stable process, which is likely to be captured by their imputation method. Second, the 20 available countries in 2000 include the main destinations in Africa (e.g. Cote d'Ivoire, South Africa) and account for 72 percent of the total immigration stock. Third, the two main missing countries (Nigeria and Ethiopia) exhibit relatively low HIV prevalence rates and have small impacts on $Z^e_{i,t}$ and $Z^i_{i,t}$.

The migration data reveal a striking fact: quantitatively, South-South migration dominates the global migrant stock, and explains one half of the world migration

stock. It is worth noticing that South-North migration is the fastest growing component of international migration, and North-South migration is negligible as all OECD countries send most of their migrants to other OECD countries. However, patterns of migration vary considerably across country pairs. We use the 1990 and 2000 migration matrices. Then, net migration flows are obtained by taking the difference: $M_{ij,t} \equiv S_{ij,t} - S_{ij,t-1}$. These net flows will be used to weight data on HIV prevalence at destination and origin as explained in equations (2) and (3). They can have positive or negative signs depending on the evolution of the stock of migrants. In our 2SLS regressions, we first predict $\widehat{S}_{ij,t}$ and $\widehat{S}_{ij,t-1}$ using the pseudo-gravity model described in Appendix B, and predict $\widehat{M}_{ij,t}$ as the difference in predicted stocks. We also extended the sample by predicting migration flows over the period 2000-2010. Our imputation methods are explained in Appendix A.

In our cross-country analysis, we use migration stock data in 1990 and 2000, and obtain one observation per country for migration net flows and HIV prevalence at destination and origin. On this basis, we estimate Eq. (1) using standard OLS and 2SLS cross-country regressions and controlling for the average level of GDP per capita over the nineties. Data on GDP per capita were taken from the *World Bank Indicators* database and are averaged to avoid losing too many observations. We also conduct the same analysis on the periods 1990-2010 and 2000-2010, based on our predicted migration stocks in 2010.

As explained above, cross-country regressions are likely to lead to inconsistent estimates. In a second step, we extend the time series dimension of our sample by proxying annual migration flows. More precisely, we annualize bilateral migration data assuming a constant annual growth rate of the migration stock over the nineties. The interpolated bilateral migration flows will then be used to weight actual annual data on HIV prevalence at destination and origin. Hence, combining primary data on HIV prevalence with annualized data on bilateral migration flows, we compute the average (or weighted) HIV prevalence levels at destination of emigrants ($Z^e_{i,t}$) and the average HIV prevalence levels at origin of immigrants ($Z^i_{i,t}$) as in (2) and (3). As bilateral migration stocks vary slowly and smoothly over time (obviously with some exceptions), our strategy looks globally reasonable and allows using panel regression techniques to solve endogeneity problems.

In this panel setting, identifying an effect of migration-weighted prevalence rates on the dynamics of HIV is possible if there is enough variability in $Z^e_{i,t}$ and $Z^i_{i,t}$. Variability comes from two sources, heterogeneous trends in bilateral migration flows and annual changes in country-specific HIV prevalence rates. Figure 2.a and 2.b depict the evolution of standardized migration-weighted prevalence rates, defined as $(Z_{i,t} - Z_{i,91}) \,/\, |Z_{i,91}|$, for all sub-Saharan African countries between 1991 and 2000. It clearly shows that trajectories of average prevalence rates at destination and origin differed a lot across countries and years. During the nineties, HIV prevalence rates increased in all countries but at very different paces, and bilateral migration trends varied a lot across country pairs although our interpolation strategy might smooth

temporal variations. Hence, the sign of variations and their relative magnitude were very heterogenous across countries; this provides a good source of identification in our fixed-effect regressions.

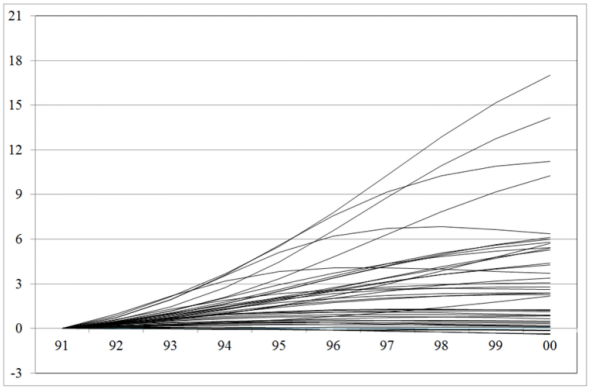
Finally, Table 1 summarizes the descriptive statistics for our main variables. We provide the sample means and standard errors calculated for the full sample of 44 sub-Saharan African countries in 1991, 2000 and 2010. For the sake of brevity, we also provide data for the 25 most infected countries in 2000. On average, the HIV prevalence rate has been multiplied by 2.5 between 1991 and 2000 (increasing from 2.83 to 7.00 percent), and has only decreased by 7 percent between 2000 and 2010 (from 7.00 to 6.49 percent). In 2000, prevalence rates range from 0.2 percent in Mauritius and Comoros to 28.6 percent in Botswana. Important changes were observed in Southern Africa (Swaziland, Lesotho, South Africa and Mozambique).

Average prevalence rates in emigration and immigration countries (expressed per 100,000 native people in Table 1) have increased in absolute value, due to the global trend in HIV. However bilateral migration flows can be negative or positive and the sign of these average rates varies across countries. In 2000, the emigration-induced rate were large in Botswana, Gabon, Côte d'Ivoire, Burkina Faso or Swaziland, and low in South Africa, Republic of Congo or Zimbabwe. As far as the immigration-induced rate is concerned, large values were reported for Liberia, Equatorial Guinea, Burkina Faso*, Mali*, Uganda, Togo*, while low levels were observed in Lesotho, Botswana, Swaziland, Namibia, Rwanda, Malawi or Zimbabwe.[10] We will take advantage of the high heterogeneity in HIV growth and prevalence rate at origin and destination to identify the migration-induced propagation mechanism.

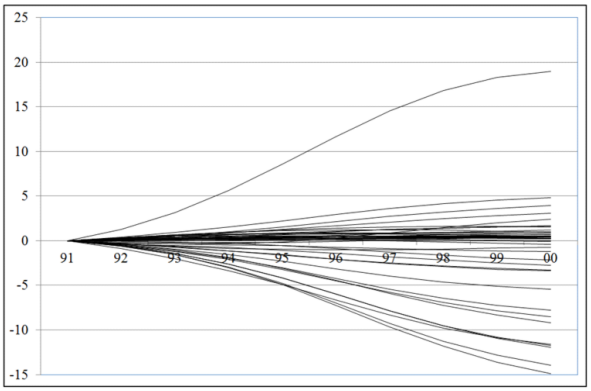
[10] A superscript * indicates that countries are not reported in Table 1, due to low levels of HIV prevalence in 2000.

Figure 2. Variability in prevalence rates ($H_{i,t}$, $Z^e_{i,t}$ and $Z^i_{i,t}$)

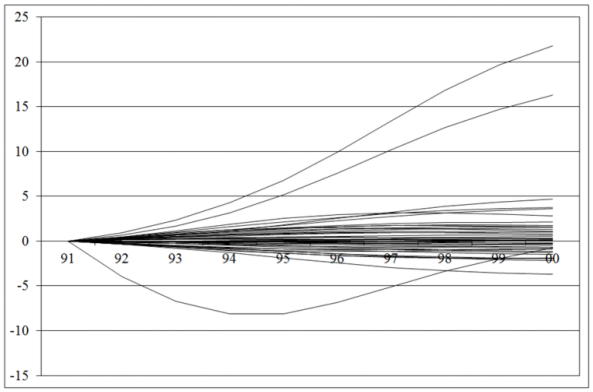
2.a. Standardized HIV prevalence rate



2.b. Standardized HIV prevalence rate at destination of emigrants



2.c. Standardized HIV prevalence rate at origin of immigrants



Note. The figure includes all sub-Saharan African countries. Prevalence rates at destination and origin are defined in Eqs (2) and (3). Values are expressed in deviation from the 1991 level, and divided by the absolute value of the 1991 level.

Table 1. Summary statistics for the years 1991, 2000 and 2010

	$H_{i,91}$	$Z^e_{i,91}$	$Z^i_{i,91}$	$H_{i,00}$	$Z^e_{i,00}$	$Z^i_{i,00}$	$H_{i,10}$	$Z^e_{i,10}$	$Z^i_{i,10}$
Sample mean	2.83	0.61	0.90	7.00	1.31	-5.35	6.49	1.37	1.03
St. error	3.66	4.34	9.15	7.79	9.18	26.98	7.46	5.78	2.76
Botswana	6.99	4.93	-4.01	28.59	28.83	-51.87	24.60	-0.07	12.44
Swaziland	2.40	0.29	-3.73	28.05	8.58	-55.77	32.31	3.41	-0.27
Zimbabwe	13.89	-1.41	-1.25	26.08	-4.09	-12.67	15.28	-14.54	-1.25
Lesotho	1.86	0.16	-9.24	23.83	1.04	-147.00	23.03	6.40	0.09
Namibia	2.57	-0.74	-3.28	18.52	-2.78	-41.82	18.54	-18.21	-0.54
Zambia	14.34	-0.87	-1.84	17.28	-1.99	-8.10	16.45	-0.87	-0.31
South Africa	1.00	-10.55	-0.16	16.49	-37.46	0.23	18.87	0.58	2.38
Malawi	6.81	-0.26	-2.49	14.72	-0.91	.16-01	12.04	-0.33	-0.15
Mozambique	0.68	0.43	-2.40	13.99	2.61	-5.34	16.67	-4.26	0.20
Cen. Afr. R.	4.91	-1.83	-1.65	11.14	-3.92	-1.03	9.92	0.14	0.00
Kenya	4.86	14.27	-0.03	8.01	9.89	-0.43	6.20	-0.28	0.10
Uganda	12.01	-6.90	6.87	7.80	-4.51	8.60	6.15	4.93	0.20
Cote d'Ivoire	5.89	10.83	1.39	7.14	17.82	1.74	4.50	1.16	8.50
Gabon	0.91	8.50	-0.45	7.10	20.88	-0.09	7.42	0.26	9.95
Tanzania	5.79	-2.12	-0.65	7.09	-3.35	-0.90	6.29	0.09	0.07
Congo, R.	6.68	-11.16	-0.45	5.85	-14.38	-1.92	4.52	-0.08	0.00
Cameroon	1.45	-0.30	0.13	5.74	-0.86	0.62	4.67	0.75	-0.31
Rwanda	8.73	0.04	-27.25	5.31	0.01	-16.52	2.78	-0.17	1.32
Angola	0.99	0.10	-0.71	3.82	0.21	-2.23	4.06	0.13	0.06
Burundi	4.18	-1.50	-3.46	3.49	-1.12	-1.97	1.75	-0.10	-0.19
Liberia	2.55	-0.14	23.37	3.46	-0.39	23.32	3.58	12.78	-0.15
Guinea-Bissau	0.46	-0.01	0.49	3.42	-0.05	1.30	4.22	1.40	-0.02
Eq. Guinea	1.43	0.50	3.98	3.31	1.70	23.02	3.66	15.15	3.19
Nigeria	0.53	0.21	0.14	3.30	0.68	0.19	3.81	0.13	0.21
Congo, DR	3.56	-0.26	-0.18	3.18	-0.29	-0.67	4.64	5.60	-1.12

Our sample includes the 44 sub-Saharan African countries. Table 1 only reports data for countries exhibiting the 25 highest HIV prevalence rates in 2000. $H_{i,t}$ = Prevalence rate per 100 people (source: UNAIDS, 2008). $Z_{i,t}$ = Prevalence rate at destination/origin per 100,000 native residents (own calculations).

4 Results

We provide four sets of results. We first provide standard cross-country results using OLS and 2SLS regressions with external instruments. Second, we use panel data and present the results obtained with the LSDV and 2SLS models. Third, we use internal instruments and re-estimate the model in level using 2SLS and in differences using GMM. Fourth, we correct for spatial correlation and describe the results of the SEM

and SAR models. In all regressions, we control for heteroskedasticity and only report robust standard errors clustered by country.

Cross-country analysis. In Table 2, we exploit cross-country data on bilateral migration and HIV prevalence to estimate Eq. (1). We first use standard OLS cross-country regressions in Columns 1-3. Then we use 2SLS regressions and instrument the norm with the predicted migration flows generated by our pseudo-gravity model (described in Appendix B). Columns 1 and 4 give the results for the period 1990-2000. In other columns, we extend the sample to 2010. It should remembered that the extension to 2010 requires using estimates of bilateral migration flows after 2000 (see Appendix A) and is more likely to suffer from mismeasurement problems. We only control for log of GDP per capita, $\ln(GDP_{i,t})$.[11]

Table 2. Cross-country regressions

Dependent variable = $\ln(1+H_{i,t}) - \ln(1+H_{i,t-n})$

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	2SLS	2SLS	2SLS
Sample (t-n,t)	1990-2000	1990-2010	2000-2010	1990-2000	1990-2010	2000-2010
$\ln(1+H_{i,t-n})$	-0.278**	-0.373***	-0.088*	-0.283***	-0.373***	-0.360***
	(0.112)	(0.060)	(0.036)	(0.110)	(0.064)	(0.865)
$\ln\left(1+Z^{e}_{i,t-n}\right)$	1.191***	1.466**	0.775*	1.170***	1.413***	1.700***
	(0.370)	(0.619)	(0.411)	(0.309)	(0.526)	(0.751)
$\ln\left(1+Z^{i}_{i,t-n}\right)$	0.089	0.051	0.021	-0.001	0.051	-0.285
	(0.083)	(0.058)	(0.026)	(0.015)	(0.053)	(0.026)
$\ln(GDP_{i,t-n})$	0.258*	0.171**	0.041	0.273***	0.171***	0.234*
	(0.134)	(0.078)	(0.039)	(0.121)	(0.063)	(0.132)
Const.	-0.553	-0.200	-0.189	0.584	-0.200	-0.408*
	(0.757)	(0.434)	(0.211)	(0.719)	(0.378)	(0.650)
# Obs.	44	88	44	44	88	44
R^2	0.323	0.364	0.212	0.300	0.364	0.230
F-Stat 1st stage	-	-	-	34.0	151.2	16.0
Cragg-Donald	-	-	-	150	96	15

Notes: $^{***}p<0.01$;$^{**}p<0.05$ and $^{*}p<0.1$. Dependent variable is the log difference of the HIV prevalence rates in t and t-n (one observation per country). OLS results are provided in Columns 1-3. 2SLS results are provided in Columns 4-6. For instrumenting the norms in 2SLS, we use the migration flows predicted by the pseudo-gravity model described in Appendix B. The level of GDP per capita and the emigration and immigration norms are constructed using the average levels observed during years t-n and t. Standard errors are clustered by country.

[11] In a previous version of this paper (see Docquier et al. 2011), we also controlled for the log of adolescent fertility rate and the log of contraceptive prevalence. Although the number of observations decreases when these controls are factored in, we obtained very similar results.

The dependent variable is defined as the log difference in the HIV prevalence rates or equivalently, the growth rate of HIV prevalence, between $t-n$ and t. All regressions include a measure of lagged prevalence rate to capture conditional convergence forces. A negative sign is always obtained for this coefficient; the model predicts a conditional convergence in HIV prevalence towards a country-specific level in the long-run.

Interestingly, the coefficient of HIV at the destination of emigrants is positive and highly significant over the period 1990-2000. This short-run elasticity is equal to 1.170 in 2SLS and 1.191 in OLS; the long-run elasticities are equal to 3.1 and 4.3, respectively. Although the data do not allow us to precisely disentangle the mechanism, this is in line with the existing literature emphasizing the role of labor migration and circulation of people in the propagation of diseases across countries.

When the sample is extended to 1990-2010, the OLS coefficient remains positive but its significance level decreases; it is only significant at 10 percent when the sample is restricted to the years 2000-2010. This might be due to different reasons: mismeasurement problems in migration flows and HIV prevalence rates are more severe after 2000, HIV prevalence rates are more stable after 2000, or the model badly explains the latter period because of omitted variables. Indeed, the 2013 UNAIDS report points out a drastic change in the dynamics of HIV after 2000. This follows substantial changes in the effectiveness of prevention policies (promotion of condoms' use, massive information campaigns for young adults, etc.) and in the access to new therapies and treatments after 2001. Absence of panel data on such policy reforms reduces the predictive power of our empirical model over the 2000-10 period. When using 2SLS regressions, the effect of HIV at destination remains significant but the R^2 of our regression falls over the period 2000-2010.

On the contrary, the coefficient of HIV at the origin of immigrants is never significantly different from zero. Consequently, cross-country regressions support a propagation mechanism through emigration, but no effect through immigration. How can we reconcile the fact that the emigration impact is significant while the effect of immigration is not? There could be several explanations. In particular, our macro analysis supports the widespread view that migrants have unprotected relations with sexual partners abroad. These partners include prostitutes in the host country, among whom infection rates are high. Hence, immigration does not induce significant changes in prevalence rates at destination (as in Wilson, 2012). However, new migrants who have been infected, propagate the virus to their origin countries through circulation, visits and/or return migration. This might explain the positive and highly significant effect of HIV at destination.

Unobserved heterogeneity. Table 3 reports the results of panel regressions for the main variables of interest. The table structure is identical to that of Table 2. What differs is that we now use annual data on HIV prevalence rates and average HIV at destination and origin of migrants. This allows us to include a full set of country and year fixed effects, which account for common time trends and unobservable characteristics of countries.

In most variants, we find out a significant and negative impact of the HIV prevalence rate of the previous period, $\ln(1+H_{i,t-1})$. Hence, the higher the HIV prevalence rate in the previous period, the lower its growth rate. The absolute value for β is lower than one. This implies that the model characterizes a stable and monotonic dynamic process through which HIV prevalence rates converge towards a country-specific long-run equilibrium (as illustrated on Figure 1). The speed of convergence is between 0.08 and 0.10, which means that it takes between 10 and 12.5 years to reach the long-run equilibrium.

Table 3. Panel regressions with fixed effects

Dependent variable = $\ln(1+H_{i,t}) - \ln(1+H_{i,t-1})$

	(1)	(2)	(3)	(4)	(5)	(6)
	LSDV	LSDV	LSDV	2SLS	2SLS	2SLS
Sample	1990-2000	1990-2010	2000-2010	1990-2000	1990-2010	2000-2010
$\ln(1+H_{i,t-1})$	-0.079***	-0.117***	-0.105**	-0.108**	-0.086**	-0.188
	(0.011)	(0.013)	(0.034)	(0.041)	(0.041)	(0.059)
$\ln\left(1+Z^e_{i,t-1}\right)$	1.329***	0.329**	0.005	0.920**	2.233*	0.280
	(0.243)	(0.135)	(0.079)	(0.375)	(2.194)	(0.323)
$\ln\left(1+Z^i_{i,t-1}\right)$	-0.092	0.000	-0.001	0.740	0.230	-0.070
	(1.845)	(0.002)	(0.003)	(0.512)	(0.723)	(0.507)
Const.	0.246***	0.219***	0.175***	0.213***	0.123*	0.213***
	(0.027)	(0.020)	(0.053)	(0.065)	(0.073)	(0.065
Country FE	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes
# Obs.	396	809	130	246	809	430
R^2	0.871	0.775	0.501	0.670	0.523	0.300
F-Stat 1st stage				500.0	130.2	638.0
Cragg-Donald				15.0	145.0	15.0
Hansen overid.				0.200	0.800	0.100

Notes: $^{***}p<0.01$;$^{**}p<0.05$ and $^{*}p<0.1$. Columns 1-3 give the results of LSDV regressions with a full set of year and country fixed effects. Columns 4-6 give the results of 2SLS regressions with a full set of year and country fixed effects. For instrumenting the norms in 2SLS, we use the annual migration flows predicted by the pseudo-gravity model described in Appendix B. Standard errors are clustered by country.

Regarding propagation mechanisms, all regressions show a positive and significant impact of the average prevalence rate at destination, $\ln\left(1+Z^e_{i,t-1}\right)$, over the period 1990-2000. With LSDV, the short-run elasticity (γ) is around 1.329 while the long-run one (γ/β) is around 16.8; with 2SLS, these elasticities decrease to 0.920 and 8.5, respectively. This effect is significant at the one percent level. The difference between LSDV and 2SLS might reflect the existence of a reverse causal link: HIV prevalence acts as a push factor for emigration. Our results strongly support the hypothesis

of conditional convergence, with long-run level depending on country characteristics and emigration patterns. As in the cross-country framework, we find no evidence of immigration-induced propagation. In all regressions, the average HIV prevalence rate at origin, $\ln\left(1+Z^i_{i,t-1}\right)$, turns out to be non significant. Finally, it is worth emphasising that the fixed effects capturing unobserved heterogeneity play a key role. They enable us to explain more than 80 percent of the variability in HIV prevalence rate over the nineties.

As in the cross-country analysis, the significance of γ falls when the sample is extended until 2010. In particular, the effect of HIV at destination is never significant over the period 2000-2010 when we account for unobserved heterogeneity. Again, this might be due to the fact that mismeasurement problems in HIV prevalence rates and migration flows are more severe after 2000. It can also be due to the fact that HIV prevalence rates exhibit much less variations after 2000, as shown in Appendix A.

Internal instruments. We now investigate in Table 4 whether our results also hold when a dynamic panel regression framework with internal instruments is used.[12] Columns 1-3 present the results obtained with a 2SLS estimation technique, with $t-2$ and $t-3$ lags of $\ln(1+H_{i,t-1})$, $\ln(1+Z^e_{i,t-1})$ and $\ln(1+Z^i_{i,t-1})$ used as instruments for their current values. In Columns 4-6, we estimate the model in differences described in equation (5) and instrument first-differenced variables using the levels in $t-3$ and $t-4$. For the model in differences, we follow Bond (2002) and use the GMM estimation method.[13] Although country fixed effects are eliminated by taking differences, we still include time fixed effects.

The main findings of the 2SLS estimations are broadly similar to those of the previous tables. We confirm a strong and significant propagation effect through emigration to infected countries over the nineties, while immigration remains insignificant. The short-run and long-run elasticities are equal to 0.994 and 7.6 (we had 0.920 and 8.5 with external instruments). No such significant effect is found for the period 2000-2010. The two necessary conditions for instrumentation are fulfilled in our regressions. Since Cragg-Donald and Stock and Yogo tests are not strictly valid in the presence of heteroskedasticity, we use the "rule of thumb" of a F-stat above 10 to test for the presence of weak instruments. In all first-stage regressions, F-stats are always far above 10 so that our instruments are not weak. They also pass the Kleinbergen-Paap's test of weak-identification test in the presence of heteroskedastic-

[12] We also tested a specification combining internal and external instruments. We instrumented the lagged dependent using its $t-2$ level (internal instrument) and instrumented the norms using predictions of our pseudo-gravity model (external instruments). We obtained very similar results. Over the period 1990-2000, the speed of convergence (β) is highly significant and equals 0.15; the effect of HIV at destination (γ) is equal to 1.047 (p-value of 0.024) and the immigration impact (δ) is small and weakly significant (it equals -0.123 with a p-value of 0.080). Our findings are very robust to the instrumentation strategy.

[13] As far as the model in difference is concerned, we also estimated it with 2SLS (all explanatory variables were instrumented with two lags) and with the *Bias-Correction LSDV* method (Bruno, 2005). We obtained similar results as in the GMM framework (see Docquier et al. 2011).

ity. Moreover, our specification is robust to the Sargan-Hansen test of joint validity of instruments.

Table 4. Panel regressions with internal instruments

Dependent variable $= \ln(1+H_{i,t}) - \ln(1+H_{i,t-1})$

	(1)	(2)	(3)	(4)	(5)	(6)
	2SLS	2SLS	2SLS	GMM	GMM	GMM
	1990-2000	1990-2010	2000-2010	1990-2000	1990-2010	2000-2010
$(\Delta)\ln(1+H_{i,t-1})$	-0.131***	-0.141***	-0.150***	0.710***	0.787***	0.750***
	(0.013)	(0.016)	(0.043)	(0.025)	(0.024)	(0.040)
$(\Delta)\ln\left(1+Z^e_{i,t-1}\right)$	0.994***	0.269*	0.017	1.210**	1.550**	0.940**
	(0.225)	(0.159)	(0.111)	(0.424)	(0.614)	(0.410)
$(\Delta)\ln\left(1+Z^i_{i,t-1}\right)$	-0.044	-0.003	-0.005	-1.611	0.370	0.170
	(1.948)	(0.005)	(0.007)	(0.411)	(0.413)	(0.380)
Const.	0.234***	0.235***	0.237***	-0.012***	0.002	0.001
	(0.027)	(0.026)	(0.069)	(0.002)	(0.003)	(0.002)
Country FE	yes	yes	yes	no	no	no
Year FE	yes	yes	yes	yes	yes	yes
# Obs	352	723	387	264	675	367
R^2	0.906	0.766	0.491	0.860	0.880	0.500
F-stat 1st stage	8373	50	34	10	23	12
KPW F-stat	1349	150	179	123	234	276
Hansen overid.	-	-	-	0.430	0.210	0.262
AR(2)	-	-	-	0.140	0.200	0.100

Notes: $^{***}p < 0.01$;$^{**}p < 0.05$ and $^{*}p < 0.1$. In columns 1-3, we provide results of 2SLS regressions with full set of year and country fixed effects. We used as instrument the second and third lags of $\ln(1+H)$ and $\ln(1+Z^e)$ as instruments. In columns 4-6, we provide results obtained with the GMM estimator and instrumented firest-differenced variables with the levels in t-3 and t-4. KPW F-stat stands for Kleinbergen-Paap Wald F-statistics; Hansen overid. stands for p-value of the Hansen overidentification test; AR(2) stands for the p-value of the AR(2) test. Standard errors are clustered by country.

As for the GMM estimations, the coefficient of the lagged term, $\triangle \ln(1+H_{i,t-1})$, is around 0.71 and highly significant. This means that the speed of convergence now increases to about 29 percent a year. It takes 3 to 4 years to reach the country-specific steady state once explanatory variables are kept constant. Previous results about migration-induced propagation effects are also comforted. The effect of immigration, $\triangle \ln\left(1+Z^i_{i,t-1}\right)$, is never significant, and the effect of emigration, $\triangle \ln\left(1+Z^e_{i,t-1}\right)$, remains positive and significant at the one percent level. The short-run elasticity increases to 1.210 and the long-run elasticiy falls to 4.2, due to the greater speed of convergence. All the tests support the fact that there is no evidence of weak instruments; and tests for second-order autocorrelation in the residuals do not reveal any

signs of additional serial correlation. All the specifications are robust to the Sargan-Hansen test of joint validity of instruments. In GMM, the effect of $\triangle \ln\left(1+Z^{e}_{i,t-1}\right)$ remains significant over the period 2000-2010 but the R^2 of this regression is much smaller.

Spatial regressions. As stated above, migration might not be the only spreading channel of HIV. To evaluate the robustness of our results, we now correct for possible spatial correlation using the spatial error (SEM) and spatial autoregressive (SAR) models. The weighting matrix of bilateral geographic distances is based on latitude and longitude data collected for the 44 sub-Saharan African countries. Columns 1-2 in Table 5 provide the results of the SEM model, assuming that spatial correlation operates through the residual term, as formalized in equation (6). In columns 3-4, we provide results for the SAR model depicted in equation (7). Our analysis is restricted to the 1990-2000 period.

Table 5. Spatial regressions (period 1990-2000)

	(1)	(2)	(3)	(4)
	SEM	SEM	SAR	SAR
Dependent	$\Delta \ln\left(1+H_{i,t}\right)$	$\Delta \ln\left(1+H_{i,t}\right)$	$\ln\left(1+H_{i,t}\right)$	$\ln\left(1+H_{i,t}\right)$
$\ln\left(1+H_{i,t-1}\right)$	-0.078***	-0.078***	0.996***	0.997***
	(0.009)	(0.010)	(0.0146)	(0.0147)
$\ln\left(1+Z^{e}_{i,t-1}\right)$	1.380***	1.330***	2.15***	2.2***
	(0.255)	(0.265)	(0.315)	(0.369)
$\ln\left(1+Z^{i}_{i,t-1}\right)$		0.072		0.9117
		(1.132)		(1.8311)
Const.	0.190	0.189	0.185	0.185
	(0.221)	(0.220)	(0.025)	(0.221)
Country FE	yes	yes	yes	yes
Year FE	yes	yes	yes	yes
# Obs.	396	396	308	308
ρ_1, ρ_2	0.120	0.118	0.1467	0.1386
	(0.079)	(0.079)	(0.097)	(0.0979)
λ			-0.1516	-0.1433
			(0.092)	(0.093)
LM test	0.106	0.108		

Notes: $^{***}p < 0.01$;$^{**}p < 0.05$ and $^{*}p < 0.1$.Spatial regressions with full set of year and country fixed effects. Columns 1-2 show results obtained with the SEM model, where ρ_1 is the spatial correlation coefficient. Columns 3-4 show regressions results obtained with the SAR model, where λ and ρ_2 are the spatial correlation coefficients. In line 'LM test', we provide the p-value of the test of no spatial error. Standard errors are clustered by country.

Qualitatively and quantitatively, results obtained with the SEM model are very similar to the LSDV ones. Both regressions point to a conditional convergence process,

a highly significant impact of average HIV prevalence at destination of emigrants, $\ln\left(1+Z^e_{i,t-1}\right)$, and an insignificant effect of average HIV prevalence at origin of immigrants, $\ln\left(1+Z^i_{i,t-1}\right)$. It is worth noticing that the spatial correlation coefficient ρ_1 is not significant, and the LM test of no spatial error is not significant. This means that we cannot reject the null hypothesis of absence of spatial correlation in the residuals: we do not need to account for spatial correlation in our preferred regressions.

The same conclusions emerge with the dynamic SAR model. The estimated spatial correlation parameters, λ and ρ_2, are never significantly different from zero. We found no evidence of dynamic or contemporaneous interactions besides the effect of emigration to countries with high infection rates. Again, it means that LSDV should be preferred to spatial correlation techniques for consistency and efficiency reasons. Emigration appears to be the main channel of transmission of HIV/AIDS between countries.

To what extent does emigration affect HIV prevalence rate? Our regressions identify a significant and very robust effect of emigration on the growth rate and level of HIV prevalence rates between 1990 and 2000. Our estimated coefficients are difficult to interpret; the reason is that we used a $\ln(1+x)$ transformation in the dependent and control variables to avoid losing countries where HIV prevalence rates are null or very small. To gauge the extent of the emigration-induced propagation mechanism, we conduct two (in-sample and out-of-sample) numerical experiments:

- First, starting from the distribution of HIV prevalence rate observed in 1990, we cut all emigration flows to all destination over the nineties (setting $\ln\left(1+Z^e_t\right) = Z^e_t = 0\ \forall t = 1990, ..., 1999$) and simulate changes in the distribution of HIV prevalence rates in 2000. From equation (1), we have:

$$\begin{aligned}\Delta_c \ln\left(1+H_{i,t}\right) &= (1+\beta)\,\Delta_c \ln\left(1+H_{i,t-1}\right) + \gamma\Delta_c \ln\left(1+Z^e_{i,t-1}\right) \\ &= (1+\beta)\,\Delta_c \ln\left(1+H_{i,t-1}\right) - \gamma \ln\left(1+Z^e_{i,t-1}\right)\end{aligned}$$

  where Δ_c means the difference between the counterfactual and the observed levels. Hence, the variation in the log of prevalence in 2000 ($\Delta_c H_{i,00}$) depends on the variation in the average prevalence rate at destination in 1999 ($\Delta_c Z^e_{i,99}$) and in all previous periods (captured by $\Delta_c H_{i,99}$). Given the time path of $Z^e_{i,t}$, we simulate the counterfactual trajectory of $\ln\left(1+H_{i,t}\right)$. Columns 3-4 in Table 6 give the counterfactual prevalence rates in 2000 and changes in the number of HIV cases as percentage of deviation from the observed level in 2000.

- Second, we simulate the long-run effect of cutting all migration flows. From equation (1), the long-run effect of cutting migration flows is given by $\gamma \ln(1+Z^e_{00})/\beta$. Columns 5-6 in Table 6 give the long-run prevalence rates and changes in the number of HIV cases as percentage of deviation from the observed levels in 2000.

Our simulation is based on parameter values obtained in Table 5, i.e $\gamma = 1.210$ and $\beta = -0.29$. Results are presented in Table 6. Column 1 reports HIV prevalence rates observed in 2000 (UNAIDS, 2008).

Table 6. No-emigration counterfactual HIV prevalence rates

	Obs 2000	Counterf. levels in 2000		Long-run counterf. levels	
	HIV rate	HIV rate[a]	HIV cases[b]	HIV rate[a]	HIV cases[b]
Regions (weighted average)					
Western Africa	0.75	2.97	-1.7	2.97	-1.7
Central Africa	2.46	4.11	+0.3	4.12	+0.4
Eastern Africa	3.91	7.32	+1.6	7.48	+2.3
Southern Africa	0.81	17.41	+4.7	18.91	+8.1
Negative migration flows					
Mauritius	0.02	0.04	+142.2%	0.06	+214.4%
Lesotho	23.8	36.5	+53.3%	47.2	+98.1%
Swaziland	28.1	32.9	+17.2%	35.9	+28.0%
Botswana	28.6	33.1	+15.9%	36.0	+25.8%
Namibia	18.5	20.9	+12.9%	22.3	+20.6%
Rwanda	5.3	6.0	+12.4%	5.8	+8.6%
Malawi	14.7	15.5	+5.2%	15.8	+7.4%
Zimbabwe	26.1	27.0	+3.6%	27.6	+5.7%
Positive migration flows					
Uganda	7.8	7.5	-4.3%	7.5	-4.0%
Togo	3.1	2.9	-4.4%	2.9	-4.5%
Mali	1.9	1.8	-5.6%	1.8	-5.7%
Eq. Guinea	3.3	3.0	-8.8%	2.9	-11.8%
Liberia	3.5	3.0	-13.2%	3.1	-11.8%
Comoros	0.01	0.02	-17.6%	0.02	-15.7%
Burkina Faso	2.3	1.7	-24.4%	1.8	-23.2%

[a] Counterfactual HIV prevalence rate in 2000 after cutting migration flows over the nineties or after 2000 (col 3) or cutting migration flows after 2000 (col 5). [b] Changes in the number of HIV cases as percentage of the observed level in 2000.

The top of the table gives results by region. Because Z_i^e can be positive or negative (depending on the sign of migration net flows), the counterfactual HIV prevalence rate can increase or decrease. Globally, net emigration flows were positive in Western Africa (i.e. bilateral migration stocks increased). In this region, cutting migration flows over the nineties would have reduced HIV prevalence rates and the number of HIV cases. By 2000, the average number of cases would have been 1.7 percent smaller. On the contrary, net emigration flows were negative in Central, Eastern and Southern Africa (i.e. bilateral migration stocks decreased). Equalizing migration flows to zero would have increased the stock of migrants and the average number

of HIV cases by 0.3 percent in Central Africa, 1.6 percent in Eastern Africa and 4.7 percent in Southern Africa. Similar qualitative patterns are obtained for the long-run experiment.

Although regional average effects are rather small, country-specific responses can be much larger. The rest of the table identifies countries in which recent migration trends have sensibly modified the trajectory of HIV prevalence rates. We only focus on sub-Saharan African countries where current emigration patterns induce a long-run change in the number of HIV cases greater than 4 percent (i.e. 15 countries out of 44). The emigration-induced propagation mechanism is important in these countries. It is less important and sometimes negligible in the other countries.

We identified a total of 20 countries in the sample where average levels of HIV at destination decreased over the nineties. This can be due to decreasing emigration flows or changes in emigrants' location (from high to low prevalence destinations). The effect on HIV prevalence has been pronounced in eight of these countries (Mauritius, Lesotho, Swaziland, Botswana, Namibia, Rwanda, Malawi and Zimbabwe), as reported in Table 6. Most of them have high prevalence rate, except Mauritius and Rwanda. From our experiments, it appears that observed decreases in emigration flows and/or HIV levels at destination have reduced HIV prevalence rates in these countries. Without such decreasing emigration (net) flows and in relative terms, long-run HIV prevalence rates would have been larger in these countries.

On the contrary, we identified a total of 24 countries where average levels of HIV at destination increased over the nineties. The effect on HIV prevalence has been pronounced in seven of these countries (Burkina Faso, Comoros, Liberia, Equatorial Guinea, Mali, Togo and Uganda), as reported in Table 6. Observed increases in emigration and/or HIV levels at destination have deteriorated HIV prevalence rates in these countries. Without such increasing emigration flows and in relative terms, HIV prevalence rates would have been lower in these countries.

5 Conclusion

This paper shows that emigration to high-prevalence destination countries tend to increase HIV prevalence rates in sub-Saharan African countries. This is compatible with the widespread view that migrants have sexual relations in their host country and propagate the virus to their origin countries through circulation, visits and/or return migration. Consequently, changes in the size and/or structure of emigration flows affect the growth rates and the levels of HIV prevalence over the nineties. The effect is very robust to the specification and the choice of the estimation method. Its magnitude varies a lot across countries. In 20 African countries (out of 44), decreasing emigration flows contributed to lower HIV risks over the nineties. In the remaining 24 countries, increasing emigration flows contributed to raise HIV prevalence. The effect was particularly strong in 15 countries, where changes in emigration flows has affected the number of HIV cases by more than 4 percent.

References

[1] Anarfi, J.K. (1993). Sexuality, migration and AIDS in Ghana - A socio-behavioural study. Health Transition Review 3, 1-22.

[2] Ateka, G.K. (2001). Factors in HIV/AIDS transmission in sub-Saharan Africa. Bulletin of the World Health Organization 79 (12), 1168-1168.

[3] Barghava, A. and F. Docquier (2008). HIV prevalence and migration of health-care staff in Africa. World Bank Economic Review 22 (2), 345-366.

[4] Bhargava, A., D.T. Jamison, L. Lau and C.J.L. Murray (2001). Modeling the effects of health on economic growth. Journal of Health Economics 20 (3), 423-440.

[5] Bazzi, S. and M. Clemens (2013). Blunt Instruments: Avoiding Common Pitfalls in Identifying the Causes of Economic Growth. American Economic Journal: Macroeconomics 5 (2), 152-86.

[6] Beine, M., F. Docquier and M. Schiff, (2013). International Migration, Transfers of Norms and Home Country Fertility. Canadian Journal of Economics 46 (4), 1406-1430.

[7] Bertoli, S. and F. Marchetta (2013). Bringing It All Back Home – Return Migration and Fertility Choices. World Development, forthcoming (http://dx.doi.org/10.1016/j.worlddev.2013.08.006).

[8] Bond, S. (2002). Dynamic panel data models: a guide to micro data methods and practice. Portuguese Economic Journal 1, 141-162.

[9] Brockerhoff, M. and A.E. Biddlecom (1999). Migration, sexual behavior and the risk of HIV in Kenya. International Migration Review 33 (4), 833-856.

[10] Brummer, D. (2002). Labour migration and HIV/AIDS in Southern Africa. IOM's Regional Office for Southern Africa: Geneva.

[11] Bruno, G.S.F. (2005). Estimation and inference in dynamic unbalanced panel data models with a small number of individuals. The Stata Journal 5 (4), 473-500.

[12] Calvo-Armengol, A., E. Patacchini and Y. Zenou (2009). Peer Effects and Social Networks in Education. Review of Economic Studies 76 (4), 1239-1267.

[13] Caselli, F., G. Esquivel and F. Lefort (1996). Reopening the convergence debate: a new look at cross-country growth empirics. Journal of Economic Growth 1 (3), 363-389.

[14] Decosas, J., F. Kane, J.K. Anarfi, K.D. Sodji and H.U. Wagner (1995). Migration and AIDS. Lancet 346 (8978), 826-828.

[15] Diamond, J. (1997). Guns, Germs, and Steel: The Fates of Human Societies. New York: W.W. Norton.

[16] Docquier, F. and H. Rapoport (2012). Globalization, brain drain and development. Journal of Economic Literature 50 (3), 681-730.

[17] Docquier, F., Ch. Vasilakis and D. Tamfutu Munsi (2011). International Migration and the Propagation of HIV in Sub-Saharan Africa. IRES Discussion Paper n. 2011-038: UCLouvain.

[18] Elhorst, J.P. (2003). Specification and Estimation of spatial panel data models. International Regional Science Review 26 (3), 826-828.

[19] Elhorst, J.P. (2010a). Spatial panel data Models. In: M.M. Fischer and A. Getis (eds.), Handbook of applied spatial analysis, 377-407, Springer: Berlin.

[20] Elhorst, J.P. (2010b). Applied Spatial econometrics: raising the bar. Spatial Economic Analysis 5 (1), 9-28.

[21] Fargues, Ph. (2007). The demographic benefit of international migration: a hypothesis and its application to Middle Eastern and North African countries. In: Ozden, C. and M. Schiff (eds), International migration, economic development and policy, World Bank and Palgrave Macmillan: Washington DC.

[22] Freyer, KJ. (2009). Trade and income-exploiting time series in geography. NBER Working Paper n. 14910, National Bureau of Economic Research.

[23] Gibson, J. and D. McKenzie (2011). Eight questions about brain drain. Journal of Economic Perspectives 25 (3), 107-128.

[24] Hatton, T. and J. Williamson (2008). Global migration and the world economy: two centuries of policy and performance. The MIT Press: Cambridge (MA) and London.

[25] Hope, K.R. (2001). Population mobility and multi-partner sex in Botswana : implications for a spread of HIV/AIDS. Journal of Reproductive Health 5 (3), 73-93.

[26] Islam, N. (1995). Growth Empirics: a Panel Data Approach. Quarterly Journal of Economics 110 (4), 1127-1170.

[27] Islam, N. (2003). What Have we Learnt From the Convergence Debate? Journal of Economic Surveys 17 (3), 309-362.

[28] Kermack, W.O. and A.G. McKendrick (1927). A Contribution to the Mathematical Theory of Epidemics. Proceedings of the Royal Society of London 115 (772), 700–721.

[29] Lee, L.F. and J. Yu (2010). A spatial dynamic panel data model with both time and individual fixed effects. Econometric Theory 26 (2), 564-597.

[30] Lodigiani, E. and S. Salomone (2012). International migration and female political empowerment. IRES Discussion Paper n. 2012-001: UCLouvain.

[31] Massey, D.S. (1988). Economic development and international migration in comparative perspectives. Population and Development Review 14 (3), 383-413.

[32] Manski, Ch. F. (1993). Identification of endogeneous social effects: the reflection problem. Review of Economic Studies 60 (3), 531-542

[33] Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. Econometrica 49 (6), 1417-1426.

[34] Ntozi, J.P.M. (1997). Widowhood, remarriage and migration during the HIV/AIDS epidemic in Uganda. Health Transition Review 7, 125-144.

[35] Oster, E. (2012). Routes of infection. Exports and HIV incidence in sub-Saharan Africa. Journal of the European Economic Association 10 (5), 1025-1058.

[36] Owen, A. and S. Wu (2007). Is Trade Good for your Health? Review of International Economics 15 (4), 660-682.

[37] Ozden, C., C. Parsons, M. Schiff and T. Walmsley (2011). The Evolution of Global Bilateral Migration 1960-2000. World Bank Economic Review 25 (1), 12-56.

[38] Pritchett, L. and L. Summers (1996). Wealthier is Healthier. Journal of Human Resources 31 (4), 841-868.

[39] Ratha, D. and W. Shaw (2007). South-South Migration and Remittances. World Bank Working Paper n. 102. The World Bank: Washington DC.

[40] Roodman, D. (2009). A note on the theme of too many instruments. Oxford Bullentin of Economics and Statistics 71 (1), 135-158.

[41] Santos Silva, J.M.C. and S. Tenreyro (2006). The log of gravity. Review of Economics and Statistics 88 (4), 641-658.

[42] Spilimbergo, A. (2009). Democracy and foreign education. American Economic Review 99 (1), 528–43.

[43] UNAIDS (2008). HIV Data. Geneva (www.unaids.org/en/hiv_data/default.asp).

[44] United Nations (2009). Trends in International Migrant Stock: The 2008 Revision. United Nations: New York.

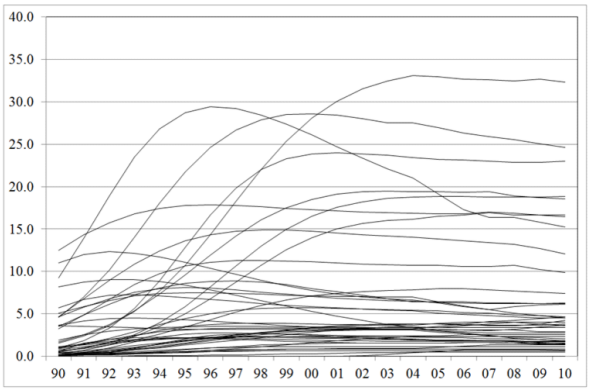
[45] Wilson, N. (2012). Economic Booms and Risky Sexual Behavior: Evidence from Zambian Copper Mining Cities. Journal of Health Economics 31 (6), 797-812.

6 Appendix

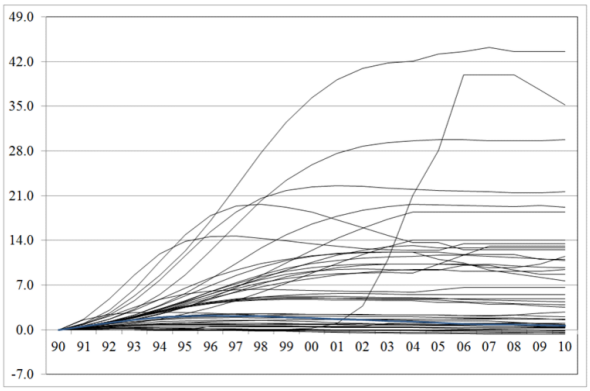
A. Data extension until 2010. As explained above, the comprehensive longitudinal data on HIV prevalence rates were revised by UNAIDS for the period for 1990–2007. Data for the period 2008-2010 can be dowloaded from the websites of UNAIDS and the World Health Organization. Figure A.1 describes the dynamics of HIV prevalence between 1990 and 2010. Figure A.1.a depicts the evolution of observed HIV prevalence rate and Figure A.1.b gives the percentage of deviation from the 1990 level.

Figure A.1. Dynamics of prevalence rates 1990-2010

A.1.a. Non standardized HIV prevalence rate



A.1.b. Standardized HIV prevalence rate



Note. The figure includes all sub-Saharan African countries. On Figure A.1.b, values are expressed in deviation from the 1990 level, and divided by the 1990 level.

Figure A.1 shows that the dynamics of HIV prevalence rates has drastically changed across decades. On the one hand, large variations were observed between

1990 and 2000. On the other hand and with a few exceptions[14], HIV prevalence rates have been relatively stable since 2000. On average, the annual growth rate in the number of HIV cases was equal to 12.3 percent in the nineties, and -0.8 percent only between 2000 and 2010.

Unfortunately, the database of Ozden et al. (2011) does not provide bilateral migration data after 2000. Computing the weighted prevalence rate in emigration and immigration countries after 2000 requires estimating bilateral migration stocks. We have used three scenarios for predicting bilateral migration stocks in 2010. The first relies on the 2010 estimates described in Ratha and Shaw (2007); the second assumes that the annual flows of migrants are identical to those observed over the nineties; and the third follows the imputation strategy used in Ozden et al (2011) for the missing countries in 1990 and 2000: we use the total stocks of immigrants in 2010 reported in the Trends in International Migrant Stock of the United Nations and assumed these stocks have the same bilateral composition as in 2000. In the paper, we only report results obtained under scenario 3. Results are similar under scenarios 1 and 2.

B. Gravity model to predict bilateral migration in 2SLS. To predict bilateral migration stocks in 1990, 2000 and 2010 ($S_{ij,t}$), we used a gravity regressions which only includes exogenous explanatory variables and fixed effects. Our gravity model writes as following:

$$S_{ij,t} = \exp\left(a_0 + a_1 \text{cmc}_{ij} + a_2 \text{cml}_{ij} + a_3 \text{same}_{ij} + a_4 \text{ldis}_{ij} + f_i + f_j + f_t\right) + u_{ij,t}$$

where cmc_{ij} is a dummy equal to one if countries i and j shared the same colonizer, cml_{ij} is equal to one if they share the same language, same_{ij} is equal to one if they were previously the same country, ldis_{ij} is the log of distance, (f_i, f_j, f_t) is a set of country and year fixed effects, and $u_{ij,t}$ is the error term. We also tried interacting the log of distance with time dummies (as in Freyer, 2009): all interaction terms were unsignificant.

The presence of a large number of zeroes in the dependent variable gives rise to econometric concerns that would yield inconsistent OLS estimates. This phenomenon is especially prevalent in migration data sets, since there is no observed or recorded migration between many country pairs due to high geographic, cultural and economic barriers. Furthermore, censuses or alternative surveying instruments are unlikely to capture small migration corridors should any sampling strategy be followed. As a result, we have zero values for 240 sub-Saharan corridors (12.5 percent of the 1,936 observations). Santos Silva and Tenreyro (2006) advocated the use of Pseudo-Poisson Maximum Likelihood (PPML) estimator that yields consistent parameter estimates even in the presence of numerous zero observations in the dependent variable.

[14]For example, the non standardized HIV prevalence rate decreased from 28 to 15 percent in Zimbabwe (a remarkable trajectory on Figure A.1.a) and increased from 0.1 to 0.7 percent in Mauritius (a remarkable trajectory on Figure A.1.b).

We thus estimated the above equation with PPML and clustered standard errors. All coefficients are significant at 5 percent, and 1 percent in most cases. We obtained a coefficient of 1.706 for cmc_{ij} (s.e.=0.831), 0.949 for cml_{ij} (s.e.=0.214), 1.875 for same_{ij} (s.e.=0.304), -1.391 for ldis_{ij} (s.e.=0.094). The coefficient of distance exceeds what has been found in studies on North-North migration, indicating that proximity is a key determinant of bilateral migration across sub-Saharan African countries. The fixed effects for the years 2000 and 2010 are equal to 0.117 (s.e.=0.067) and 0.255 (s.e.=0.111), respectively. The R^2 of this regression amounts to 0.896 but should be taken with caution in a PPML framework. We use the estimated equation to predict bilateral migration stocks in 1990, 2000 and 2010 ($\widehat{S}_{ij,t}$), and then predict migration flows by differencing the stocks ($\widehat{M}_{ij,t} = \widehat{S}_{ij,t} - \widehat{S}_{ij,t-1}$). In the panel setting, we annualize predicted bilateral migration data assuming a constant annual growth rate of the predicted migration stock.

C. Robustness to alternative specifications. Our preferred model is a standard β-convergence specification which features the annual log-change in HIV prevalence as the dependent variable and three main explanatory variables: past level of HIV prevalence, average level of HIV prevalence in destination countries of native emigrants, average level of HIV prevalence in origin countries of foreign immigrants. We used a specification with $\ln(1+x)$ to avoid losing observations with $x \simeq 0$, i.e. to be consistent with countries where HIV prevalence rates (domestic, at destination or at origin) are null or very small. In this Appendix, we present the results obtained with two alternative specifications.

First, epidemiological models have been used to model the progress of an epidemic in a large population. It divides the population in three compartments, individuals who are susceptible to the disease (S), infected individuals (I) and those who have recovered (R). In the SIR model, the dynamics of the prevalence rate, ΔI_t, is characterized by the following differential equation: $\beta S_t I_t - R_t$. In the case of HIV, we eliminate R_t so that the propagation of the epidemic is proportional to the product of lagged rates of infected ($H_{i,t-1}/100$) and non-infected individuals ($1-H_{i,t-1}/100$). We introduce the average prevalence rate at destination of emigrants as another potential control and estimate the model with or without fixed effects:

$$\Delta \ln(1+H_{i,t}) = \alpha + \beta \frac{H_{i,t-n}}{100}\left(1 - \frac{H_{i,t-n}}{100}\right) + \gamma \ln(1+Z^e_{i,t-n}) + \alpha_i + \alpha_t + \varepsilon_{i,t}$$

In Table A1, Column 2 gives our panel estimation results and Column 1 gives the results of a cross-country regression without fixed effects and for the period 1990-2000. Our coefficient of interest, γ, is never significant. However the SIR coefficient β is also unsignificiant, which means that the standard epidemiological model is rejected by the data.

Another feature of our preferred model is that it uses a specification with $\ln(1+x)$ to be consistent with countries where prevalence rates very small ($H \simeq 0$ or $Z \simeq 0$). Here we test a specification with $\ln(x)$ for all variables. Coefficients can be interpreted

as elasticities. Results are presented in Column 3 for the cross-country OLS regression (pooling 1990-2000 and 2000-2010 data) and in Column 4 for the panel regression with fixed effects. In both cases, the speed of convergence is larger than with the $\ln(1+H)$ specification, and the average prevalence rate at destination of emigrants remain highly significant. In the panel regression, the coefficient obtained for HIV at destination is unsurprisingly smaller than with the $\ln(1+Z)$ transformation; this is due to the fact that $\ln(Z)$ exhibits greater variations than $\ln(1+Z)$. In addition, the R^2 is smaller than that obtained with the $\ln(1+x)$ transformation (compare column 4 with column 1 in Table 3).

Table A1. Results obtained under alternative specifications

	(1)	(2)	(3)	(4)
	Cross-country	Panel	Cross-country	Panel
Dependent	$\Delta \ln(1+H_{i,t})$	$\Delta \ln(1+H_{i,t})$	$\Delta \ln(H_{i,t})$	$\Delta \ln(H_{i,t})$
$\frac{H_{i,t-n}}{100}\left(1-\frac{H_{i,t-n}}{100}\right)$	-0.0016	0.0003	-	-
	(0.0056)	(0.0007)	-	-
$\ln(H_{i,t-n})$	-	-	-0.5276***	-0.1780***
	-	-	(0.0510)	(0.0350)
$\ln\left(1+Z^e_{i,t-1}\right)$	0.0250	0.4632		
	(0.0288)	(0.6611)		
$\ln\left(Z^e_{i,t-1}\right)$			2.9578***	0.1164***
			(0.4304)	(0.0418)
$\ln(\text{GDP}_{i,90})$	0.1891	-	0.2980***	-
	(0.1296)	-	(0.0877)	-
Const.	-0.5446	0.0489***	-0.7679	0.2856***
	(0.7729)	(0.0070)	(0.5130)	(0.0401)
Country FE	no	yes	no	yes
Year FE	no	yes	no	yes
# Obs.	44	396	77	728
R^2	0.3655	0.8732	0.6512	0.8525

Notes: $^{***}p < 0.01;^{**}p < 0.05$ and $^{*}p < 0.1$. Dependent variable is the log difference of the HIV prevalence rates in t and t-n. OLS results are provided in Columns 1 and 3. LSDV results are provided in Columns 2 and 4. Standard errors are clustered by country.

"Sur quoi la fondera-t-il l'économie du monde qu'il veut gouverner? Sera-ce sur le caprice de chaque particulier? Quelle confusion! Sera-ce sur la justice? Il l'ignore."

Pascal

FERDI

Créée en 2003, la **Fondation pour les études et recherches sur le développement international** vise à favoriser la compréhension du développement économique international et des politiques qui l'influencent.

Contact
www.ferdi.fr
contact@ferdi.fr
+33 (0)4 73 17 75 30